Genome-Wide DNA Methylation Profiling in Dietary Intervention Studies: a User's Perspective

Clarissa Gerhauser · Katharina Heilmann · Maria Pudenz

Abstract Within the past decade, epigenetic mechanisms and their modulation by dietary agents have gained major interest in the cancer prevention and nutrition community. Gene expression is epigenetically regulated by DNA methylation, histone tail modifications, and non-coding (micro) RNAs. Given the fact that epigenetic aberrations are reversible and represent potentially initiating events in the development of diseases, they have been identified as promising new targets for prevention strategies. Evidence is accumulating that dietary cancer chemopreventive agents from various sources, including green tea, soy, turmeric, broccoli, and other fruit and vegetables, can modulate DNA methylation, at least in vitro. To facilitate in vivo studies with focus on genome-wide modulation of DNA methylation, we here give an overview on current affinity enrichment- and bisulfite treatment-based methodologies for methylation profiling that might be useful for rodent models and human intervention studies. We also summarize genome-wide methylome analyses performed with dietary agents in vitro and in vivo and conclude with some practical considerations for the design of future dietary intervention studies.

Keywords Epigenetic · Reduced representation bisulfite sequencing (RRBS) · Whole genome bisulfite sequencing (WGBS) · Illumina 450k array · Methylated CpG immunoprecipitation (MCIp) · Next-generation sequencing (NGS)

C. Gerhauser · K. Heilmann · M. Pudenz
Epigenomics and Cancer Risk Factors, Workgroup Cancer Chemoprevention and Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany

C. Gerhauser (✉)
Epigenomics and Cancer Risk Factors, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
e-mail: c.gerhauser@dkfz.de

Introduction

Dietary Agents Targeting the Epigenome

Within the past decade, epigenetic mechanisms and their modulation by dietary agents have gained major interest in the cancer prevention and nutrition community. Numerous in vitro and selected in vivo analyses indicate that aberrant epigenetic programming during disease progression may be prevented and even reversed by dietary agents. Phytochemicals from various dietary sources, including green tea, soy, fruit, and berries such as black raspberries, cruciferous vegetables, turmeric, onions, cashew nuts, and others, were shown to directly target enzymatic activities or modulate expression of enzymes involved in epigenetic gene regulation, including DNA methyltransferases (DNMTs) and histone-modifying enzymes such as histone acetyltransferases, deacetylases, methyltransferases, and demethylases that modulate chromatin accessibility. Many phytochemicals were also shown to alter expression of non-coding (micro) RNAs in cell culture, adding to their potential to epigenetically regulate gene expression. Research is accumulating that these activities might contribute to chemopreventive efficacy by affecting signal transduction cascades mediated by nuclear receptors and transcription factors such as NF-κB, cell proliferation and cell cycle progression, cellular differentiation, DNA repair, apoptosis induction, cell motility, metastasis formation, and cellular senescence (reviewed in [1–11]). If true, these agents could be of significant value in cancer prevention. So far, evidence for in vivo epigenetic activities in animal models and human pilot studies and its relevance for

chemopreventive efficacy is limited. Careful documentation of how these agents impact epigenetic programming in tissues is critical for understanding their impact on the epigenome in cancer prevention.

DNA Methylation

DNA methylation is one of the best investigated mechanisms of epigenetic gene regulation [12, 13]. The transfer of methyl groups to DNA is catalyzed by the DNMT family of enzymes. In mammals, DNA methylation mainly occurs at the 5-position of cytosine (C) in the context of CpG dinucleotides, generating 5-methylcytosine (5mC). The current human genome build contains more than 28 M CpGs dinucleotides. Interestingly, CpG sites are not evenly distributed in the genome: there is an accumulation of CpGs (CpG-dense regions, so-called CpG islands or CGIs) in promoter regions of genes, whereas intra and intergenic regions are characterized by a lower density of CpGs. In healthy tissue, promoter CGIs are normally unmethylated, allowing active gene transcription, whereas non-promoter CpGs are highly methylated, thus limiting DNA accessibility and contributing to genomic stability [14]. As tissues age, an increasing number of genetic loci become silenced by DNA methylation, a process that is likely to be exacerbated by poor health and nutrition. Gene silencing often continues to expand through the carcinogenic pathway, with a range of critical growth-regulatory and tumor suppressor genes targeted in cancers. Global loss of methylation (hypomethylation), especially at repetitive sequences, and hypermethylation of CGIs in promoter regions are among the most important epigenetic changes to occur in cancer cells and thought to be involved in the etiology of cancer. In contrast to the irreversible inactivation of tumor suppressor genes by genetic alterations, genes silenced by epigenetic modifications are still intact and can be reactivated [14].

So far, most of the studies investigating the influence of dietary agents on aberrant DNA methylation have been performed in (cancer) cell culture and focused on only few selected candidate genes. Consequently, at present, it is largely unknown whether promoter demethylation and reactivation of genes silenced by DNA methylation is a random effect accompanying unspecific inhibition or reduced expression of DNMTs or whether there are targeted mechanisms underlying these activities.

Genome-Wide Methylation Profiling

With the advancement and increased affordability of array- and next-generation sequencing (NGS)-based technologies, we now have tools at hand for epigenetic analyses at a genome-wide level. Methodologies for genome-wide methylation profiling are either enrichment-based by affinity of

methylated DNA to methyl-binding proteins or antibodies against 5mC or rely on the quantitative determination of DNA methylation levels after bisulfite-treatment of DNA (overview in Table 1). A comprehensive description of the techniques is beyond the scope of this perspective. Readers are referred to recent articles and reviews that have benchmarked the methods in detail [15–18, 19••, 20–22]. A good comparison of principles and limitations/sources for bias is given in ref. [23••].

Both enrichment approaches have been coupled with DNA microarray hybridization (Chip) and/or massively parallel sequencing (Seq) for identification of thousands of genomic regions differentially methylated between tumor and normal tissues at a genome-wide scale.

Changes in DNA methylation after short- or intermediate-term dietary intervention are expected to be small. Therefore, selection of methodology for the unbiased detection of small genome-wide DNA methylation changes in nutritional studies remains a challenge. We will here describe our practical experiences and give examples of ongoing studies.

Affinity Enrichment-Based Methods

One of the strategies to reduce complexity in whole-genome methylation analyses is to enrich for highly methylated regions. Differentially methylated regions (DMRs) are then identified by comparison of two distinct samples (for example, tumor vs. normal). Enrichment is achieved either by methylated DNA immunoprecipitation (MeDIP) with monoclonal antibodies against 5mC [24] or by methods based on affinity capture of methylated CpGs with family members of the methyl-CpG binding domain (MBD)-based proteins (collectively termed as MBDCap) [25]. Several MBDCap methods have been developed that differ in the MBD protein used for enrichment. The MethylCap (methylated DNA capture) assay uses the MBD domain of MeCP2 [26], MCIp (methylated CpG immunoprecipitation) [27], and MiGS (MBD-isolated genome sequencing) [28] employ MBD2 protein, and MIRA (methylated-CpG island recovery assay) uses a complex of MBD-based proteins MBD2 and MBD3L1 with enhanced affinity for methylated CpGs compared to MBD2 alone [29, 30].

There are differences between MeDIP and MBDCap methods that influence the obtained results (see ref. [21]): the anti-5mC antibody used for MeDIP captures DNA fragments containing one or more methylated cytosines that are then eluted in one fraction. In contrast, MBD proteins bind with increasing affinity to multiple methylated CpG dinucleotides in close proximity. One can take advantage of this fact by serial elution of methylated DNA fragments with increasing salt concentrations. Downstream analyses of multiple fractions provide an overview of methylation changes at regions with increasing CpG density as exemplified in ref. [26], at the expense of higher costs. Alternatively, one can focus on

Table 1 Overview of technologies for genome-wide methylome analyses

Detection method	Principle	Data acquisition	Species	DNA requirements	Coverage preference	Coverage no. of regions/ CpGs	Resolution (bp)	Bioinformatic requirements	Costs
Affinity-based									
MeDIP	Affinity to antibody against 5mC	Array	Array: lim.[a]	1.5 µg	Low-intermediate CpG density	Dep. on array[d]	~150	++	++
		Seq	Seq: any[b]					+++	+
MBDCap methods: MethylCap	Affinity to MBD proteins: MeCP2	Array	Array: lim.[a]	0.2–3 µg	High CpG density[c]	Dep. on array[d]	~150	++	++
MCIp/MiGS MIRA	MBD2 MBD2/MBD3L1	Seq	Seq: any[b]			~5M[e]		+++	+
Bisulfite conversion-based									
(T)-WGBS		Seq	Any[a]	~10 ng	None	27.5 M	1	+++	+++
RRBS	DNA digestion with MspI, size selection	Seq	Any[a]	100 ng	Intermediate-high CpG density	~3 M	1	+++	+
Illumina 27k	Binding to specific probes	Array	Human	250 ng	CGIs	27,578	1	+	+
Illumina 450k	Binding to specific probes	Array	Human	250 ng	Dep. on array design	485,000	1	+	+

MeDIP "methylated DNA immunoprecipitation" [24], *MBDCap methods* affinity capture with methyl-CpG binding domain (MBD)-based proteins [25], *MethylCap* "methylated DNA capture" [26], *MCIp* "methylated-CpG-immunoprecipitation" [27], *MiGS* "MBD-isolated genome sequencing" [28], *MIRA* "methylated-CpG island recovery assay" [29, 30], *RRBS* "reduced representation bisulfite sequencing" [50], *(T)-WGBS* "(tagmentation-based) whole genome bisulfite sequencing" [49, 51], *Seq* massive parallel sequencing

[a] Limited to species with commercially available arrays

[b] Species with complete reference genome

[c] Depends on assay design and salt gradient for elution

[d] Dependent on species and array design. For example, CGI arrays from Agilent cover 27.700 CGIs with 237.000 probes for human samples and 15.300 CGIs with 88.000 probes for mouse samples

[e] Dependent on protocol and required level of coverage

the highest affinity fraction eluted with a high salt (HS) concentration to enrich for CGIs, or perform "single-fraction (SF) elution" without fractionation according to CpG density, as described previously [21, 25, 31]. In general, MeDIP was found to preferentially enrich low to intermediated CpG dense regions, whereas MBDCap methods, depending on the elution protocol, could be biased for CpG-dense regions [22, 25].

MCIp-Chip Analysis of Human Breast Cancer

Both MeDIP and MBDCap methods have been used in combination with promoter, CGI or tiling arrays, or NGS. Array-based detection has the advantage of relatively straightforward bioinformatical analysis, but is biased by the selection of genomic regions covered by the array. We used MCIp-based enrichment by HS elution of CpG-dense, highly methylated DNA fragments from breast cancer and normal breast tissue in combination with comparative hybridization to CGI arrays to discover a series of hypermethylated genes as novel potential biomarkers of low-grade breast cancer [32]. Identification of regions with most consisted gain in methylation was achieved by generating histograms of probes that met certain cutoff criteria (details in [32]) (Fig. 1a). Significant hypermethylation in tumor tissue was confirmed and validated in independent sample sets by quantitative mass spectrometry-based EpiTyper MassArray technology [33], with methlylation differences of 20 to >60 % between tumor and normal tissues.

In general, detection of differential methylation by MBDCap methods is biased by copy number alterations [21]. For example, in regions affected by loss of heterozygosity, only one allele is enriched, and the region is then detected as false-positive hypomethylated DMR. Often, we could not validate such "loss of methylation" by EpiTyper MassArray, which does not discriminate between alleles. On the other hand, methylation levels of hypermethylated DMRs were >80 % concordant between both technologies. Consequently, we limited our analyses to hypermethylated regions.

MCIp-Seq in the C3(1) Transgenic Murine Breast Cancer Model

We currently employ robot-assisted MCIp-enrichment by HS elution of highly methylated CpG-dense regions coupled with NGS (Heilmann et al., in preparation) to characterize changes in DNA methylation during the progression of breast cancer in the transgenic *C3(1)SV40TAg* (C3(1)) mouse strain [34]. For the analysis of methylation kinetics, mammary glands and tumors of transgenic C3(1) mice (TG) were collected at 4-week intervals over a period of 24 weeks (three animals per age group). Mammary glands of age-matched wild-type (WT) littermates that do not show tumor formation were analyzed for comparison. Quality-controlled raw sequencing reads were aligned to the mouse reference genome using Burrow-

Wheeler-Alignment ([35], summary of computational methods and software tools for methylation analyses in ref. [36••]). Depending on the degree of enrichment, NGS will identify different numbers of fragments covering a specific region (reads) in a sample. In contrast to the readout by CGI arrays, NGS allows genome-wide detection of enriched fragments, and we achieved an overall coverage of about 3.4–6.7 M CpGs. We used the HOMER tool (hypergeometric

Fig. 1 Examples of affinity enrichment-based genome-wide methylation ▶ analyses (MCIp-Chip, MCIp-Seq). **a** MCIp-Chip analysis of human breast cancer, with KCTD8 as an example (from ref. [32]). Highly methylated DNA fragments from 10 human breast cancer and 10 normal breast tissues were enriched by MCIp with high salt (HS) elution. Unmatched pairs of tumor and normal samples were co-hybridized to Agilent CpG island (CGI) arrays for the detection of differentially methylated regions (DMRs). The genomic position (in *blue/black*), location of the corresponding CGI (in *green*), and probes covered on the array (in *grey*) are depicted. The number of arrays for which a given probe was positive (from 0 to 10) is represented by the height of the corresponding *red* bar. For comparison, WGBS data tracks for HCC1954 breast cancer and HMEC human mammary epithelial cells (in *yellow*) were derived from the UCSC genome browser [83, 84] DNA methylation track hubs [85, 86]. Each *vertical line* represents one CpG site; the height indicates the level of methylation from 0 to 100 %. The CGI located at the KCTD8 promoter is highly methylated in HCC1954, but unmethylated in HMEC, and dependent on the probe, was positive in up to 10/10 arrays. **b** Genomic distribution of hyper- and hypomethylated DMRs identified by MCIp-Seq in tumors derived from the C3(1) transgenic mouse model in comparison to mammary glands of wild-type controls. More than 90 % of the DMRs are located outside of core promoter regions (including 5′UTR, promoters, and CGIs). **c** Kinetics of hypomethylation at a CGI (location in *green*) overlapping an exon of the Daxx gene (in blue) in the C3(1) mouse model. Highly methylated DNA fragments were enriched with HS elution and subjected to next-generation sequencing (NGS). Read counts (range 0–100 reads, normalized to 10 M reads) at specific genomic positions are depicted for groups of transgenic (TG, *red*) and wild-type (WT, *blue*) animals covering an age range of 4–24 weeks. Noteworthy is the excellent visual uniformity of MCIp-enriched DNA fragments in all groups. Comparison with WBGS tracks for murine embryonic stem cells (ESC), placenta and uterus (in *yellow*) indicates that MCIp-enrichment with HS elution is limited to fragments with high CpG density (*narrow lines* in WGBS tracks) and high methylation. **d** Confirmation by quantitative EpiTyper MassArray analyses of consistent high methylation in WT mice vs. gradual loss of methylation in TG animals at the intragenic/exonic Daxx CGI. Median and range in each group are indicated by *black lines*. **e** MCIp-Seq analyses to identify methylation changes in mammary glands of ovariectomized Wistar rats exposed to a soy isoflavone-enriched diet (IRD) vs. isoflavone-depleted diet (IDD) for 17 days after ovariectomy. The *bottom track* shows the percentage of G (guanine)+C (cytosine) bases in 5-base windows (in *black*, range 30–70 % with a *horizontal line* at 50 %). As in Fig. 1c, HS elutes CpG dense DNA fragments (*upper two green tracks*, range 0–30 reads, normalized to 10 M reads). Note that the CGI in the promoter region of SFRP1 (secreted frizzled related protein 1) is not methylated in normal mammary glands and therefore not enriched. For comparison, we performed single fraction (SF) elution to enrich DNA fragments with intermediate-high CpG density (*lower two red tracks*, range 0–10 reads normalized to 10 M reads), as seen by the increased number of intra- and intergenic peaks that are not enriched by HS elution. By SF elution, more CpGs are covered at least once, but to reach saturation comparable to HS elution, an overall higher number of reads per sample is required
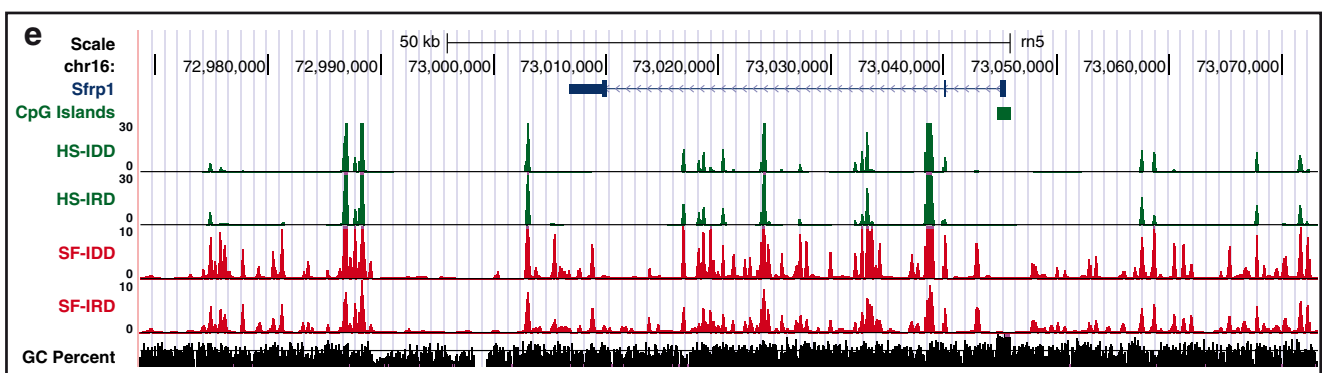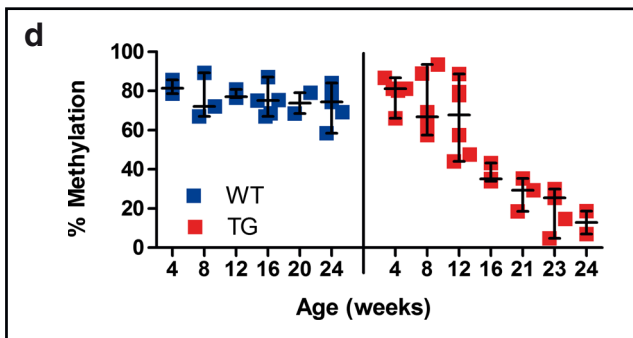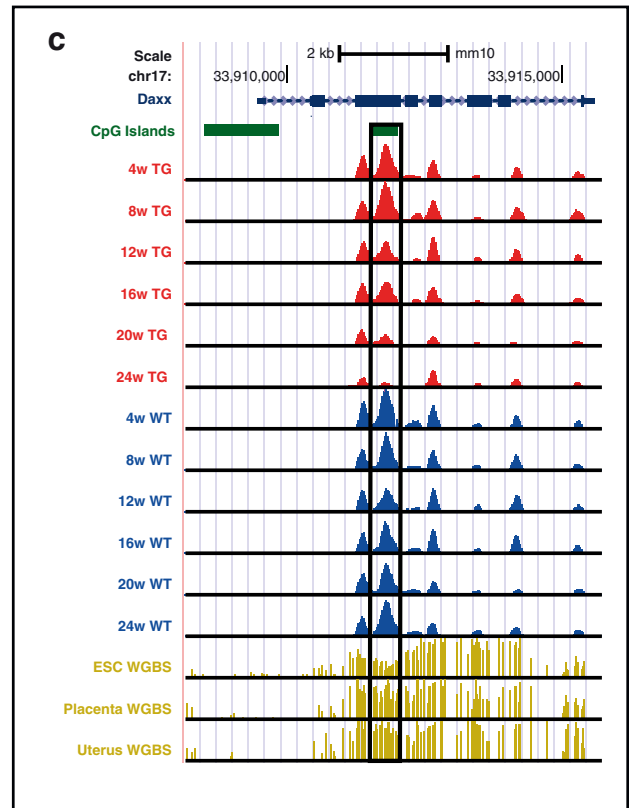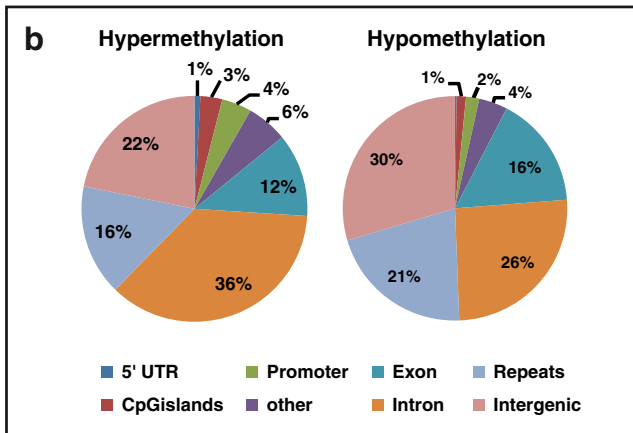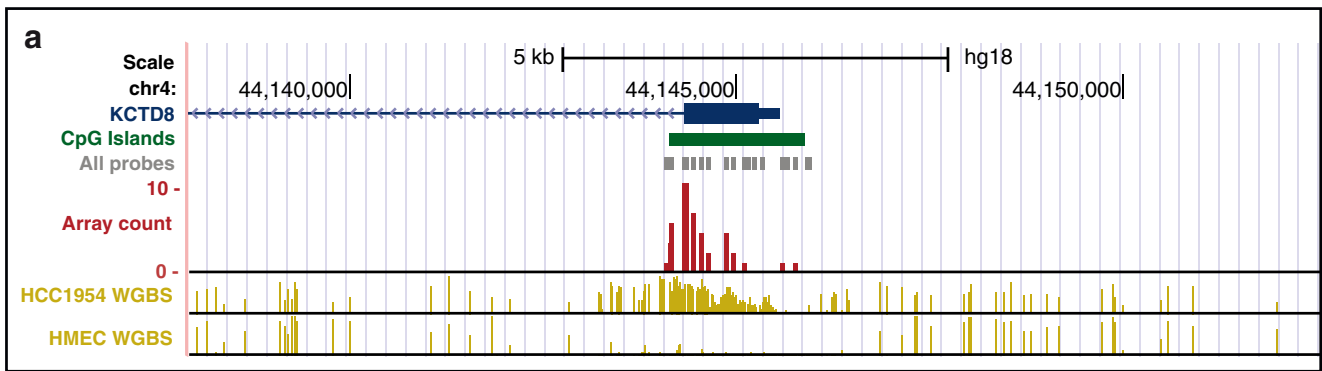
optimization of motif enrichment [37]) to calculate differences in read frequencies between TG and WT animals per age group, taking defined parameters such as fold change (FC), p value, and false discovery rate (FDR) into consideration. As an example, at the age of 24 weeks, we detected around 9.000 hyper and 9.500 hypomethylated DMRs based on a fourfold difference in read counts between TG and WT. Notable, more than 90 % of all DMRs (Fig. 1b) were located outside of core promoter regions that contribute to gene regulation (including 5′UTR, promoters, and CGIs). Serre et al. reported similar findings for MiGS analyses [28]. For validation, we selected, among others, a DMR overlapping with an exonic CGI of Daxx (death-domain-associated protein), a histone chaperone that facilitates chromatin assembly [38]. MCIp-Seq revealed a gradual loss of methylation in TG mice over time (Fig. 1c), which was associated with significant upregulation of gene expression (data from [39]). EpiTyper MassArray analyses confirmed stable methylation at around 80 % in WT mice and significant hypomethylation in TG animals from 80 % at 4 weeks to around 13 % median methylation at 24 weeks (Fig. 1d). Since we selected DMRs with gradual changes in methylation over time and integrated data on methylation changes with changes in gene expression, different from the human study, in the C3(1) study, both hyper and hypomethylated DMRs could be validated by EpiTyper MassArray analyses with >90 % concordance.

As mentioned above, the majority of DMRs were located outside of core promoter regions. We are currently just beginning to understand the role of DMRs in intronic, exonic, and intergenic regions for general and cell-type-specific gene regulation. They have been postulated to overlap with enhancers or transcription factor binding sites [40, 41]. For dietary intervention studies in healthy volunteers, methylation changes at these regions might be of interest, as promoter CGIs are generally not methylated in normal tissues and hence, only gain in methylation would be detectable.

MCIp-Seq After Dietary Soy Intervention in Healthy Wistar Rats

MCIp-Seq is used in an ongoing study on methylation changes in mammary glands of Wistar rats as part of the German Research Foundation (DGF)-funded IsoCross project, in cooperation with partners from the German Sports University in Cologne (Pudenz et al., in preparation). One of the aims in this project is to investigate the influence soy isoflavones on estrogen sensitivity of mammary glands [42]. Isoflavones are phyto-estrogenic plant compounds and could therefore attenuate physiological changes associated with hormone deprivation [43]. A group of female rats was subjected to ovariectomy at the age of 80 days. During the period of hormonal decline after ovariectomy, rats received a diet enriched with a soy extract equivalent of about 400 ppm isoflavones

(isoflavone-rich diet, IRD) or an isoflavone-depleted diet (IDD) for comparison. The study was terminated at day 97, and DNA from mammary glands was processed by MCIp-Seq as described for the C3(1) study. We first focused on the analyses of CpG-dense fragments with high DNA methylation obtained by HS elution. Bioinformatic comparison of both groups revealed that about 3100 regions were hypo and 2000 regions hypermethylated in the IRD vs. the IDD group, with fourfold difference in read counts between both groups, and at least 10 reads per region in one group as selection criteria (Fig. 1e, upper two lanes). Again, more than 90 % of the DMRs were located outside of core promoter regions.

Different from the C3(1) study, the identified DMRs were difficult to confirm by quantitative EpiTyper MassArray analyses. We often obtained non-significant median methylation differences below 10 % between IRD- and IDD-treated groups and observed high inter-individual variation of animals within one group. The latter might be attributed to the fact that the Wistar rat strain is an outbred stain, and individual response to ovariectomy and isoflavone intervention might vary substantially. The discordance between MCIp enrichment and EpiTyper MassArray results might result from the fact that MCIp with HS elution will enrich CpG-dense, highly methylated DNA fragments, whereas CpG-dense regions with lower DNA methylation as well as highly methylated DNA with low CpG density are removed before sequencing (also, see ref. [31]). Therefore, by setting a high detection threshold, small absolute differences between groups will be amplified. Discordance might also result from the fact that we analyzed whole genomic DNA from a part of one mammary gland per animal. Mammary glands are composed of multiple cell types, for example, ductal epithelial cells, adipocytes, stromal cells, immunological cells, and others. A difference in MCIp-Seq read counts between groups might reflect enrichment of highly methylated DNA derived from a subpopulation of cells. In subsequent quantitative EpiTyper MassArray analyses of the bulk of DNA derived from all cells in one sample, methylation differences will be "diluted" by the contribution of all cell types to an average methylation level. Differences in read counts in MCIp-Seq experiments might also reflect differences in cell composition between samples rather than alterations in methylation levels. This bias can be avoided by using pre-selected cell populations (for example, by cell sorting based on surface markers using magnetic beads or flow cytometry or by laser capture microdissection of regions of interest).

It has been proposed that highly methylated CGIs might be more resistant to demethylation than regions with intermediate levels of methylation [44]. Therefore, we performed another MCIp-enrichment from the same samples by SF elution with a reduced salt concentration (also, see refs. [21, 31]) to include DNA fragments with intermediate and high

CpG-densities and/or methylation levels (Fig. 1e, lower two lanes). Obviously, we obtained higher coverage of the genome compared to the HS elution, but the distribution between various genomic locations did not considerably change. However, it should be noted that substantially higher numbers of mapped reads are required to reach comparable saturation (Pearson correlation of two random partitions of the sequenced sample, indicative of reproducible coverage of the reference genome [45, 46]) as with HS elution. Therefore, we had to pool sequencing reads from several samples to reach sufficient coverage. Still, aligned read counts were lower than for HS elution (as reflected by the different y-axis scale for HS and SF elution in Fig. 1e). Keeping identical selection criteria (fourfold difference and at least 10 reads per region in one group) we detected about 300 hypo and 1000 hypermethylated DMRs between the IRD and the control group. Validation by EpiTyper MassArray is currently ongoing; therefore, we cannot comment yet whether this approach might be a recommended strategy for intervention studies.

In conclusion, to enhance the sensitivity of detecting small methylation differences at genome-wide level and to increase the coverage of genomic regions with low CpG-density and methylation, SF elution seems to be the preferred method. However, to reach sufficient saturation and coverage, a higher number of sequencing reads is required, increasing the overall costs of the analyses. In any case, enrichment-based methods provide only relative or indirect information on DNA methylation levels that have to be validated by independent quantitative methods such as EpiTyper MassArray or pyrosequencing [47].

Quantitative Bisulfite-Treatment-Based Methods

Sodium bisulfite (BS) conversion of genomic DNA is the gold standard for DNA methylation analysis to differentiate and detect unmethylated versus methylated cytosines [48]. BS treatment of single-stranded DNA results in the preferential chemical deamination of unmethylated cytosine residues to uracil, whereas the deamination of 5mC to thymine is very slow. In subsequent PCR reactions, all uracils (from unmethylated Cs) are amplified as thymines, whereas only 5mCs are amplified as cytosines, allowing discrimination of unmethylated and methylated Cs at single CpG resolution. Incomplete BS conversion results in false-positive detection of Cs as 5mCs. Since after BS treatment opposite DNA strands are no longer complementary, BS conversion-based methods permit strand-specific methylation analysis.

WGBS and RRBS

Genome-wide methods that utilize the advantages of BS treatment include conventional and tagmentation-based whole genome bisulfite sequencing (WGBS, T-WGBS) and reduced representation bisulfite sequencing (RRBS) [49–51]. WGBS provides single nucleotide methylation information for about 95 % of all CpGs in a genome (examplary, WGBS reference tracks are depicted in Figs. 1 and 2). T-WGBS uses an alternative protocol for WGBS library preparation, based on the enzymatic activity of a transposase that simultaneously fragments the DNA and tags the fragments with adapters. This procedure makes intermediate cleanups between different library preparation steps largely unnecessary; thus, the amount of DNA input can be reduced to 10–30 ng [51].

For both WGBS and T-WGBS, an extended number of reads (obtained from about three lanes on an Illumina HiSeq 2000 flow cell with paired-end sequencing of 100 bp) is required to sufficiently cover the entire genome. Therefore, these methods are extremely costly and due to the large amount of data generated demand extended bioinformatical expertise. In mammalian genomes, only a small fraction of all Cs is methylated (about 3–6 %). As a consequence, after BS treatment the genomic sequence is reduced mainly to three bases T, A, and G, hampering mapability of the obtained reads to the reference genome [15]. To overcome these complications, specific bioinformatic tools are being developed to process datasets from WGBS (overview in ref. [36••]). Different from array-based technologies, sequencing-based methods can be applied to any species as long as a reference genome is available.

RRBS uses the same NGS strategy as WGBS [19••]. The fraction of the genome to be sequenced is reduced by digesting genomic DNA with restriction endonucleases that are specific for CpG containing motifs, in combination with fragment size selection [52]. About 1–3 M CpGs are covered by RRBS, with enhanced coverage of regions with moderate to high CpG density, including CGIs, promoters, and enhancers. Since RRBS fragments the DNA at specific restriction sites, the analyzed fragments for a given species are relatively constant, thus increasing the utility for comparative DNA methylation profiling [19••, 53].

In cooperation with Christoph Bock (CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna), we have established a RRBS library preparation pipeline and subjected first samples derived from the IsoCross project for sequencing. Based on our experience with MCIp-Seq followed by EpiTyper MassArray analyses ("MCIp-Seq After Dietary Soy Intervention in Healthy Wistar Rats"), we expect methylation changes to be small. Validated software tools such as RnBeads [54] are available for RRBS data mining that present quantitative methylation data at single CpG site, CGI and promoter level, allow pair- or groupwise statistical comparison of intervention groups for efficient detection of DMRs, generate graphical reports of results, and provide an overview of enriched gene ontologies representing the DMRs [55]. Our expectations are that this methodology will provide more reliable genome-wide methylation profiles from dietary intervention studies, as it is not
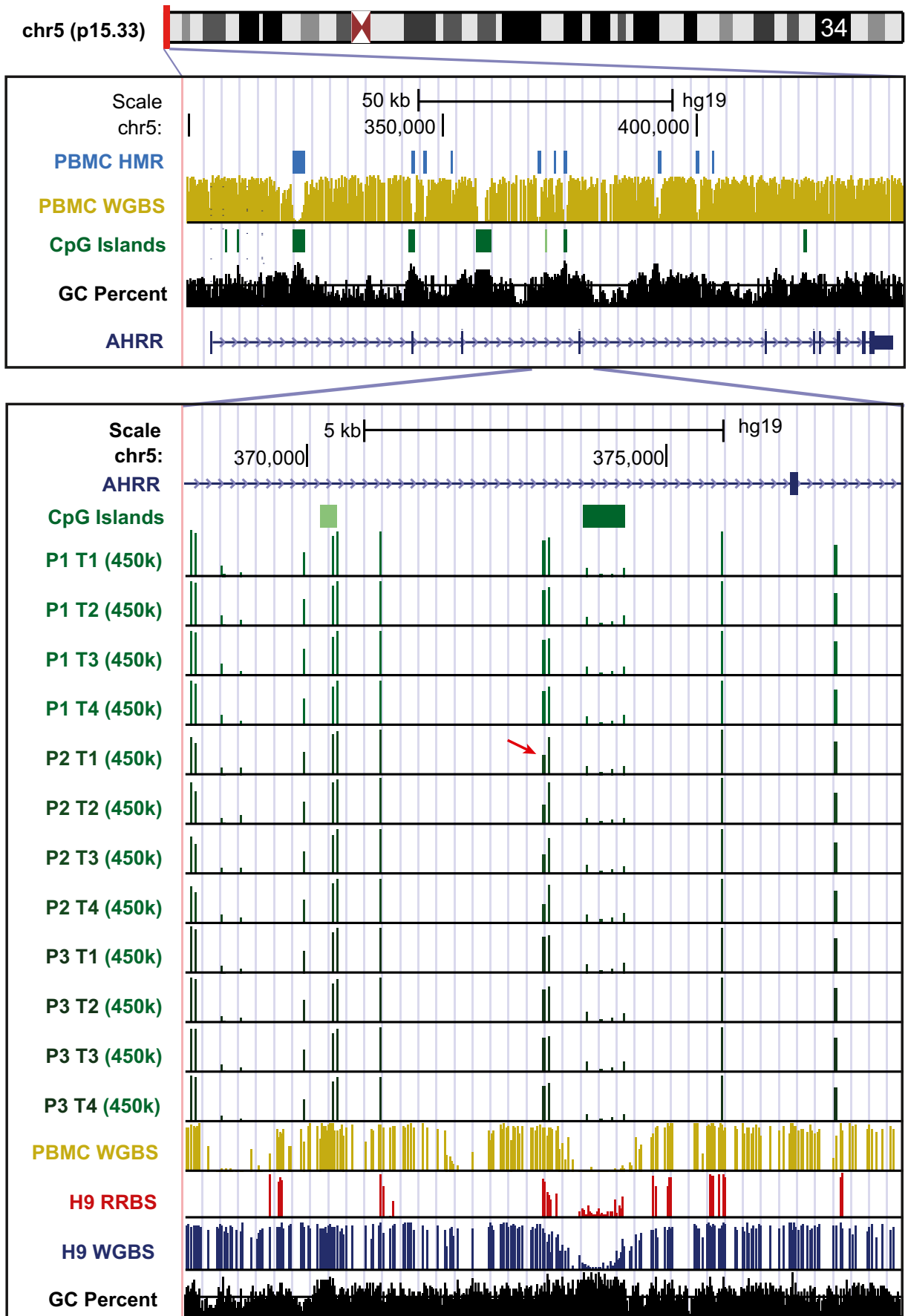
Fig. 2 Examples of bisulfite treatment-based genome-wide methylation analyses (WGBS, Illumina 450k, RRBS). Depicted is the genomic location of the AHRR gene (aryl hydrocarbon receptor repressor), which harbors a CpG site (cg23576855, indicated by *red arrow*) with methylation levels associated with smoking status [60]. UCSC genome browser panels show (from *top* to *bottom*): chromosomal location, graph scale, hypomethylated regions (HMR, marked with *blue bars*) in peripheral blood mononuclear cells (PBMC), PBMC WBGS information (in *yellow*) derived from the UCSG genome browser DNA methylation track, with each *vertical line* representing one CpG site and the height indicating the level of methylation from 0 to 100 %, location of CGIs (indicated in *green*), GC percentage (in *black*, scale 30–70 %, with a *horizontal line* at 50 %), AHRR gene locus in *blue/black*. The region of interest (marked by a *red box*) is enlarged below in order to visualize details with higher resolution. The lower panel depicts (from *top* to *bottom*): the enlarged intragenic region of the AHRR locus with scale, chromosomal region, AHRR gene locus in *blue/black*, location of CGIs (in *green*), followed by 12 tracks (in *green*) of methylation data from 450k analysis of PBMC DNA derived from three young healthy smokers (P1, P2, P3) participating in a pilot intervention study over a period of 6 weeks. Time points (T1–T4) indicate: *T1* before and *T2* after placebo intervention (10 days), 20 days washout period, *T3* before and *T4* after broccoli intervention (for 10 days). Each *vertical line* represents the location of one CpG site covered on the 450k array; the height indicates methylation levels from 0 to 100 %. WGBS information for PBMC DNA (in *yellow*) is followed by two tracks for the human embryonic stem cell line H9. H9 RRBS (in *red*) indicates quantitative methylation information (from 0 to 100 %) derived from RRBS analyses (each *vertical line* represents one CpG site), with H9 WGBS information (in *blue*) for comparison. The bottom track gives an overview of the G+C content (range 30–70 %, with a *horizontal line* at 50 %). Note that (i) 450k analyses recapitulate WBGS information at reduced coverage, (ii) probes overlapping with SNPs (such as the one indicated by the *red arrow*) should be excluded from genome-wide analyses across individuals, unless the influence of SNPs on methylation levels is part of the research question, (iii) 450k and RRBS provide comparable datasets

enrichment-based and provides a direct readout of quantitative methylation levels for biostatistical comparison of samples. Sample requirements in the range of 100 ng genomic DNA are lower than for MBDCap-Seq methods, thus facilitating analysis of preselected subpopulations of cells. When several samples are multiplexed for sequencing, costs for both technologies are comparable. Depending on the quality of sequencing libraries, 1–3 M CpG sites can be covered per sample. It should however be kept in mind that the analyzable CpG sites are not necessarily identical in all samples of one experiment (different from array analyses with defined sites) and consequently the number of overlapping CpG sites that meet certain quality criteria (e.g., coverage) might drop substantially.

Illumina 27k and 450k Beadchip Arrays

As an alternative to sequencing-based methods for samples from human intervention studies, Illumina Infinium Human Methylation 27k and 450k Bead Chip arrays provide quantitative methylation data of defined CpG sites at a genome-wide level.

First introduced in 2008, the Illumina 27k array platform covers 27,578 CpG sites located in CGI associated with 14,475 annotated genes. BS-treated DNA is hybridized to a set of bead-bound probes, one designed against the methylated and one against the unmethylated C at each locus. After hybridization, single base extension with labeled nucleotides incorporates a fluorescent label for detection, thus adding another level of specificity. The Illumina 450k array is a further development of the Illumina 27k platform. Over 480,000 probes cover 99 % of the annotated genes with an average of 17 CpGs per gene. The 450k platform covers 96 % of CGI with additional coverage of CGI shores and CGIs outside coding regions [56]. More than 41 % (>197,000) sites are located in intergenic regions (bioinformatically predicted enhancers, DNase I hypersensitive sites, and validated DMRs) [19••]. Since its introduction in early 2011, 450k technology is now widely used in international large scale epigenomic profiling projects such as "The Cancer Genome Atlas" [57], providing thousands of reference epigenome data sets for normal tissues and multiple cancer types. The 450k array is based on two different assay types to interrogate methylation levels. Type I assays are equivalent to 27k technology and overlap to 90 % with sites covered on 27k arrays, whereas type II assays rely on only one probe per site and discriminate the methylation status by the labeled nucleotide incorporated. Since both types of assay perform differently, care has to be taken to properly normalize methylation levels for comparative results. Also, a fraction of the probes overlaps with single nucleotide polymorphisms (SNPs) thus introducing an analytical bias. Nevertheless, 450k technology allows fast and cost-efficient genome-wide analysis of DNA methylation, requires relatively small amounts of input DNA (0.25–0.5 µg), is compatible with DNA isolated from archival samples including FFPE tissues, and can be processed in a high-throughput manner [19••].

In a small pilot study in cooperation with the University of Milan (Italy), we used 450k technology to interrogate DNA methylation changes in peripheral blood mononuclear cells (PBMCs) of three young healthy smokers who participated in a short 10-day intervention with steam-cooked broccoli (250 g/day) to modulate smoking-associated oxidative stress. The study was placebo-controlled, with a 20-day washout period between placebo- and broccoli-diet interventions. Blood was taken before and after both intervention periods, providing four DNA samples per volunteer over a period of about 6 weeks (study details in [58]). All 12 DNA samples were analyzed on one 450k bead chip, thus avoiding problems associated with batch affects. Over the study period, PBMC methylation was very stable (example in Fig. 2). Significant changes in DNA methylation in samples from before and after broccoli intervention larger than 5 % were infrequent and

Table 2 Genome-wide methylation profiling with cancer preventive agents

Chemopreventive agent	Cell lines—animal model	Dose and duration	Detection method	Comments	Ref.
In vitro					
Curcumin from turmeric	HCT116, HT29, and RKO colon cancer cells	7.5–10 μM, 6 or 240 days	Illumina 27k	• No short-term effects • No changes in global methylation • 814–3051 differentially methylated sites with methylation changes >10 %, mostly in CGIs with intermediate methylation levels	[44]
Genistein from soy	ARCaP-E and ARCaP-M prostate cancer cells	20 μM, 6 days	Illumina 27k	• No significant changes in methylation >20 %	[69]
G2535 (purified soy extract)	C4-2B and LNCaP prostate cancer cells	Equivalent to 20 μM genistein, 5 days	Illumina 450k	• Genome-wide data not provided • Demethylation and upregulation of microRNAs miR-29a and miR-1256	[87]
Sulforaphane, 3,3′-diindolylmethane derived from broccoli	PrEC, LNCaP, PC3 prostate (cancer) cells	15 μM, 3 days	MeDIP-Chip	• Widespread promoter hypo- and hypermethylation in all three cell lines • Gene targets of both compounds are distinct between the three cell lines, but largely overlap within one cell line • Methylation changes in LNCaP vs. normal PrEC are normalized by both compounds	[74]
In vivo					
Green tea extract	Transgenic adenoma of the mouse prostate (TRAMP) model and wild-type (WT) mice	0.3 % in drinking water, from 4 to 24 weeks of age; 0.1, 0.3, and 0.6 % in drinking water, from 6 to 18 weeks of age	LC-MS, HELP-assay	• No significant changes in global methylation • Relative detection of genome-wide changes with HELP assay, changes not consistent between wild-type and TRAMP samples	[81]

G2535 purified soy extract, contains 70.5 % genistein, 26.3 % daidzein, 0.3 % glycitein; *LC-MS* liquid chromatography-electrospray ionization quadrupole mass spectrometry; *HELP assay* HpaII tiny fragment enrichment by ligation-mediated PCR assay

demonstrate similar broad and complex effects on DNA methylation profiles in in vivo models.

Green tea polyphenols (GTP) with (-)-epigallocatechin gallate (EGCG) as the major catechin belong to the best characterized cancer chemopreventive agents. Tea and tea constituents act by a broad spectrum of anti-carcinogenic activities and were reported first to affect DNA methylation in 2003 (reviewed in [3, 5]). EGCG and GTP demonstrate convincing cancer preventive efficacy in animal models (review in [75, 76]), including the "Transgenic Adenocarcinoma of the Mouse Prostate" (TRAMP) model [77–80]. In a study published in 2009, Morey-Kinney et al. analyzed the impact of green tea extract intervention on prostate carcinogenesis and genome-wide methylation changes in TRAMP mice [81]. Unexpectedly, different from previous reports, GTP (0.3 % in drinking water) did not prevent the development of prostate tumors. Wild-type (WT) and TRAMP mice were exposed to water or water supplemented with 0.3 % GTP starting from 4 weeks of age up to 24 weeks. Global methylation (5mC) levels were quantified by liquid chromatography coupled with mass spectrometry (LC-MS) in the gut, liver, and prostate of WT mice and prostate tumors of TRAMP mice. About 3–4.5 % of all Cs were methylated, and 5mC levels did not differ between intervention groups, except for a significant 0.5 % reduction upon GTP treatment in livers of WT animals at 12 weeks of age. Quantitative methylation analysis by EpiTyper MassArray of four selected candidate genes known to become hypermethylated in prostate tumors of TRAMP mice revealed age-specific changes, but no influence of the GTP intervention. Global methylation levels were also not affected in TRAMP mice by dose-dependent intervention with 0.1, 0.3, and 0.6 % GTPs from 6 to 18 weeks of age. In order to determine methylation changes at a genome-wide level, the HELP (HpaII tiny fragment enrichment by ligation mediated PCR) assay was applied to one selected TRAMP and WT prostate per group (at age 24 weeks), with and without 0.3 % GTP treatment. This assay detects relative methylation changes at about 1 M loci in the mouse genome. GTP induced both hyper- and hypomethylation when compared within each mouse strain; however, the changes were not concordant between WT and TRAMP samples. Unfortunately, data were not confirmed by quantitative methods, and with only one sample analyzed per group, this study lacked statistical power to identify significant methylation changes.

Summary and Conclusions

The present perspective aims to give an overview of current methodology for assessing DNA methylation at genome-wide scale and to summarize practical experience with applications for methylome profiling in two breast cancer-related projects and dietary intervention studies. The tool set for genome-wide methylation analysis ranges from enrichment-based methods with array- or sequencing-based detection to WGBS that allows species-independent detection of nearly all CpGs in a genome at single CpG resolution, at the expense of high costs.

Selection of the most suitable methodology strongly depends on the research question and the expected degree in methylation changes. Cancer development is associated with continuous changes in DNA methylation at CGIs and core promoter regions. Each of the described methods will accurately detect these extensive methylation differences, which can be as high as 60–80 % when comparing tumor samples and normal tissues. Regions outside of core regulatory regions recently revealed to be important for gene regulation, including intra- and intergenic enhancer regions and areas affected by larger scale methylation changes, such as partially methylated domains and DNA methylation valleys (for example, in [82]) are only partly covered on available array platforms. Sequencing-based methods including WGBS would therefore be the preferred methodology, but are bioinformatically more demanding than array-based methods. Enrichment-based methods are sensitive, but strongly depend on the protocol used for elution of enriched DNA fragments, and the envisaged saturation and coverage will determine required reads. Also, they can be biased by CpG density and copy number alterations.

From what we know so far, short-term dietary intervention in healthy subjects will induce maximally around 10 % change in methylation (when analyzing mixed cell populations). Therefore, unequivocal detection at a genome-wide scale can be challenging. Bisulfite conversion-based methods provide quantitative readout and have low sample requirements, therefore allowing pre-selection of most interesting cell populations to enhance detection sensitivity. For human intervention studies, Illumina 450k arrays offer a good compromise of costs, bioinformatic demands and genomic coverage and have been shown to provide reliable data highly correlated with results obtained with alternative quantitative methods. Information on predefined CpG sites allows straightforward comparison between samples within one study, but eventually also between studies and research groups. Also of interest, large datasets for various normal tissues and cancer entities are publicly available for comparison, for example, from TCGA [57]. For rodent studies or cross-species comparison, RRBS provides equivalent data as 450k analyses, with high coverage of genomic regions with intermediate-high CpG density. With respect to efficient use of resources, one might consider performing preliminary WGBS analyses with a limited number of samples (minimum $n=3$) to identify DMRs with the highest possible resolution and sensitivity, and follow up on larger sample sets with quantitative methods such as EpiTyper MassArray or pyrosequencing. Evidently, however, such small-scale analysis looses statistical power to detect methylation differences and is biased by the selected sample set.

In order to obtain informative results from intervention studies, not only the choice of methodology but also the study design should be carefully considered. DNA methylation is intricately involved in developmental processes. It has been speculated that dietary cancer preventive agents might function by induction of epigenetic reprogramming during development. Therefore, studies with interventions starting early in life, covering developmentally critical time windows, might provide more meaningful outcome than studies with interventions in adult animals or adult human subjects. When early intervention is not feasible, interventions for extended periods of time will have larger impact on the epigenome than short-term treatments for few weeks.

With a series of complementary technologies for genome-wide methylation analysis now at hand, future research will have to focus on integration of effects on various epigenomic mechanisms with gene expression and the link to disease outcome to identify best strategies for dietary intervention targeting the epigenome.

Compliance with Ethics Guidelines

Conflict of Interest Clarissa Gerhauser, Katharina Heilmann, and Maria Pudenz declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent The C3(1) mouse study has been approved by the regional board in Karlsruhe (Germany) with the reference number 35-9185.82/A-15/08. Details and permission for the soy intervention study in rats are described in ref. [42]. Details and ethical approval for the human broccoli intervention study are described in ref. [58].

References

Papers of particular interest, published recently, have been highlighted as:
•• Of major importance

1. Li Y, Tollefsbol TO. Impact on DNA methylation in cancer prevention and therapy by bioactive dietary components. Curr Med Chem. 2010;17(20):2141–51. doi:BSP/CMC/E-Pub/ 128.

2. Vanden Berghe W. Epigenetic impact of dietary polyphenols in cancer chemoprevention: lifelong remodeling of our epigenomes. Pharmacol Res. 2012;65(6):565–76. doi:10.1016/j.phrs.2012.03.007.

3. Gerhauser C. Epigenetics, (poly) phenolics and cancer prevention. In: Romani A, Lattanzio V, Quideau S, editors. Recent Advances in Polyphenol Research, Vol. 4: Wiley-Blackwell, UK; 2014 (in press).

4. Gerhauser C. Cancer chemoprevention and nutriepigenetics: state of the art and future challenges. Top Curr Chem. 2013;329:73–132. doi:10.1007/128_2012_360.

5. Huang J, Plass C, Gerhauser C. Cancer chemoprevention by targeting the epigenome. Curr Drug Targets. 2011;12(13):1925–56. doi:BSP/CDT/E-Pub/00214.

6. Link A, Balaguer F, Goel A. Cancer chemoprevention by dietary polyphenols: promising role for epigenetics. Biochem Pharmacol. 2010;80(12):1771–92. doi:10.1016/j.bcp.2010.06.036.

7. vel Szic KS, Ndlovu MN, Haegeman G, Vanden Berghe W. Nature or nurture: let food be your epigenetic medicine in chronic inflammatory disorders. Biochem Pharmacol. 2010;80(12):1816–32. doi:10.1016/j.bcp.2010.07.029.

8. Meeran SM, Ahmed A, Tollefsbol TO. Epigenetic targets of bioactive dietary components for cancer prevention and therapy. Clin Epigenetics. 2010;1(3–4):101–16. doi:10.1007/s13148-010-0011-5.

9. Rajendran P, Ho E, Williams DE, Dashwood RH. Dietary phytochemicals, HDAC inhibition, and DNA damage/repair defects in cancer cells. Clin Epigenetics. 2011;3(1):4. doi:10.1186/1868-7083-3-4.

10. Malireddy S, Kotha SR, Secor JD, Gurney TO, Abbott JL, Maulik G, et al. Phytochemical antioxidants modulate mammalian cellular epigenome: implications in health and disease. Antioxid Redox Signal. 2012;17(2):327–39. doi:10.1089/ars.2012.4600.

11. Supic G, Jagodic M, Magic Z. Epigenetics: a new link between nutrition and cancer. Nutr Cancer. 2013;65(6):781–92. doi:10.1080/01635581.2013.805794.

12. Berdasco M, Esteller M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev Cell. 2010;19(5):698–711. doi:10.1016/j.devcel.2010.10.005.

13. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13(7):484–92. doi:10.1038/nrg3230.

14. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet. 2007;8(4):286–98. doi:10.1038/nrg2005.

15. Wong E, Wei CL. Genome-wide distribution of DNA methylation at single-nucleotide resolution. Prog Mol Biol Transl Sci. 2011;101:459–77. doi:10.1016/B978-0-12-387685-0.00015-9.

16. Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol. 2010;28(10):1097–105. doi:10.1038/nbt.1682.

17. Meaburn E, Schulz R. Next generation sequencing in epigenetics: insights and challenges. Semin Cell Dev Biol. 2012;23(2):192–9. doi:10.1016/j.semcdb.2011.10.010.

18. Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, et al. Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol. 2010;28(10):1106–14. doi:10.1038/nbt.1681.

19. •• Stirzaker C, Taberlay PC, Statham AL, Clark SJ. Mining cancer methylomes: prospects and challenges. Trends Genet. 2014;30(2):75–84. doi:10.1016/j.tig.2013.11.004. *Excellent comparison of techniques available for genome-wide DNA methylation analysis.*

20. Mansego ML, Milagro FI, Campion J, Martinez JA. Techniques of DNA methylation analysis with nutritional applications. J Nutrigenet Nutrigenomics. 2013;6(2):83–96. doi:10.1159/000350749.

21. Robinson MD, Statham AL, Speed TP, Clark SJ. Protocol matters: which methylome are you actually studying? Epigenomics. 2010;2(4):587–98. doi:10.2217/epi.10.36.

22. Li N, Ye M, Li Y, Yan Z, Butcher LM, Sun J, et al. Whole genome DNA methylation analysis based on high throughput sequencing technology. Methods. 2010;52(3):203–12. doi:10.1016/j.ymeth.2010.04.009.

23. •• Laird PW. Principles and challenges of genomewide DNA methylation analysis. Nat Rev Genet. 2010;11(3):191–203. doi:10.1038/nrg2732. *This review expertly summarizes the advantages and disadvantates of techniques for genome-wide DNA methylation analyses and points out potential limitations.*

24. Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, et al. Methylome analysis using MeDIP-seq with low DNA concentrations. Nat Protoc. 2012;7(4):617–36. doi:10.1038/nprot.2012.012.

25. Nair SS, Coolen MW, Stirzaker C, Song JZ, Statham AL, Strbenac D, et al. Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. Epigenetics. 2011;6(1):34–44. doi:10.4161/epi.6.1.13313.

26. Brinkman AB, Simmer F, Ma K, Kaan A, Zhu J, Stunnenberg HG. Whole-genome DNA methylation profiling using MethylCap-seq. Methods. 2010;52(3):232–6. doi:10.1016/j.ymeth.2010.06.012.

27. Gebhard C, Schwarzfischer L, Pham TH, Schilling E, Klug M, Andreesen R, et al. Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. Cancer Res. 2006;66(12):6118–28. doi:10.1158/0008-5472.CAN-06-0376.

28. Serre D, Lee BH, Ting AH. MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res. 2010;38(2):391–9. doi:10.1093/nar/gkp992.

29. Rauch TA, Pfeifer GP. DNA methylation profiling using the methylated-CpG island recovery assay (MIRA). Methods. 2010;52(3):213–7. doi:10.1016/j.ymeth.2010.03.004.

30. Rauch T, Li H, Wu X, Pfeifer GP. MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. Cancer Res. 2006;66(16):7939–47. doi:10.1158/0008-5472.CAN-06-1888.

31. Robinson MD, Stirzaker C, Statham AL, Coolen MW, Song JZ, Nair SS, et al. Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation. Genome Res. 2010;20(12):1719–29. doi:10.1101/gr.110601.110.

32. Faryna M, Konermann C, Aulmann S, Bermejo JL, Brugger M, Diederichs S, et al. Genome-wide methylation screen in low-grade breast cancer identifies novel epigenetically altered genes as potential biomarkers for tumor diagnosis. FASEB J. 2012;26(12):4937–50. doi:10.1096/fj.12-209502.

33. Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A. 2005;102(44):15785–90. doi:10.1073/pnas.0507816102.

34. Maroulakou IG, Anver M, Garrett L, Green JE. Prostate and mammary adenocarcinoma in transgenic mice carrying a rat C3(1) simian virus 40 large tumor antigen fusion gene. Proc Natl Acad Sci U S A. 1994;91(23):11236–40.

35. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60. doi:10.1093/bioinformatics/btp324.

36. •• Bock C. Analysing and interpreting DNA methylation data. Nat Rev Genet. 2012;13(10):705–19. doi:10.1038/nrg3273. *This review describes the workflow for analysing and interpreting genome-scale DNA methylation data, gives an overview of software tools for the analysis and interpretation of DNA methylation data. It also describes common problems associated with DNA methylome analyses and gives potential solutions.*

37. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576–89. doi:10.1016/j.molcel.2010.05.004.

38. DeNizio JE, Elsasser SJ, Black BE. DAXX co-folds with H3.3/H4 using high local stability conferred by the H3.3 variant recognition residues. Nucleic Acids Res. 2014;42(7):4318–31. doi:10.1093/nar/gku090.

39. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol. 2007;8(5):R76. doi:10.1186/gb-2007-8-5-r76.

40. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507(7493):455–61. doi:10.1038/nature12787.

41. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;500(7463):477–81. doi:10.1038/nature12433.

42. Blei T, Soukup S, Schmalbach K, Pudenz M, Möller F, Egert B, et al. Dose-dependent effects of isoflavone exposure during early lifetime on the rat mammary gland: Studies on estrogen sensitivity, isoflavone metablosim and DNA methylation. Mol Nutr Food Res. 2014. doi:10.1002/mnfr.201400480.

43. Pudenz M, Roth K, Gerhauser C. Impact of Soy isoflavones on the epigenome in cancer prevention. Nutrients. 2014;6(10):4218–72. doi:10.3390/nu6104218.

44. Link A, Balaguer F, Shen Y, Lozano JJ, Leung HC, Boland CR, et al. Curcumin modulates DNA methylation in colorectal cancer cells. PLoS ONE. 2013;8(2):e57709. doi:10.1371/journal.pone.0057709.

45. Trimarchi MP, Murphy M, Frankhouser D, Rodriguez BA, Curfman J, Marcucci G, et al. Enrichment-based DNA methylation analysis using next-generation sequencing: sample exclusion, estimating changes in global methylation, and the contribution of replicate lanes. BMC Genomics. 2012;13 Suppl 8:S6. doi:10.1186/1471-2164-13-S8-S6.

46. Chavez L, Jozefczuk J, Grimm C, Dietrich J, Timmermann B, Lehrach H, et al. Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage. Genome Res. 2010;20(10):1441–50. doi:10.1101/gr.110114.110.

47. Tost J, Gut IG. DNA methylation analysis by pyrosequencing. Nat Protoc. 2007;2(9):2265–75. doi:10.1038/nprot.2007.314.

48. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A. 1992;89(5):1827–31.

49. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315–22. doi:10.1038/nature08514.

50. Gu H, Bock C, Mikkelsen TS, Jager N, Smith ZD, Tomazou E, et al. Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. Nat Methods. 2010;7(2):133–6. doi:10.1038/nmeth.1414.

51. Wang Q, Gu L, Adey A, Radlwimmer B, Wang W, Hovestadt V, et al. Tagmentation-based whole-genome bisulfite sequencing. Nat Protoc. 2013;8(10):2022–32. doi:10.1038/nprot.2013.118.

52. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. Preparation of reduced representation bisulfite sequencing libraries

for genome-scale DNA methylation profiling. Nat Protoc. 2011;6(4):468–81. doi:10.1038/nprot.2010.190.

53. Smith ZD, Gu H, Bock C, Gnirke A, Meissner A. High-throughput bisulfite sequencing in mammalian genomes. Methods. 2009;48(3):226–32. doi:10.1016/j.ymeth.2009.05.003.

54. MPI for Informatics. RnBeads [updated 8.8.2014]. Available from: http://rnbeads.mpi-inf.mpg.de/.

55. Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. Nat Methods. 2014. doi:10.1038/nmeth.3115.

56. Morris T, Lowe R. Report on the Infinium 450 k methylation array analysis workshop: April 20, 2012 UCL, London, UK. Epigenetics. 2012;7(8):961–2. doi:10.4161/epi.20941.

57. TCGA. The cancer genome atlas available from: http://cancergenome.nih.gov.

58. Riso P, Martini D, Moller P, Loft S, Bonacina G, Moro M, et al. DNA damage and repair activity after broccoli intake in young healthy smokers. Mutagenesis. 2010;25(6):595–602. doi:10.1093/mutage/geq045.

59. Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. Environ Health Perspect. 2014;122(2):131–7. doi:10.1289/ehp.1306937.

60. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. 2013;22(5):843–51. doi:10.1093/hmg/dds488.

61. Brennan K, Garcia-Closas M, Orr N, Fletcher O, Jones M, Ashworth A, et al. Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. Cancer Res. 2012;72(9):2304–13. doi:10.1158/0008-5472.CAN-11-3157.

62. Wang X, Zhu H, Snieder H, Su S, Munn D, Harshfield G, et al. Obesity related methylation changes in DNA of peripheral blood leukocytes. BMC Med. 2010;8:87. doi:10.1186/1741-7015-8-87.

63. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinforma. 2012;13: 86. doi:10.1186/1471-2105-13-86.

64. Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS ONE. 2012;7(7):e41361. doi:10.1371/journal.pone.0041361.

65. Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G, et al. Heterogeneity in white blood cells has potential to confound DNA methylation measurements. PLoS ONE. 2012;7(10):e46705. doi:10.1371/journal.pone.0046705.

66. Gupta SC, Sung B, Kim JH, Prasad S, Li S, Aggarwal BB. Multitargeting by turmeric, the golden spice: from kitchen to clinic. Mol Nutr Food Res. 2013;57(9):1510–28. doi:10.1002/mnfr.201100741.

67. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. J Clin Invest. 2009;119(6):1420–8. doi:10.1172/JCI39104.

68. Banerjee S, Li Y, Wang Z, Sarkar FH. Multi-targeted therapy of cancer by genistein. Cancer Lett. 2008;269(2):226–42. doi:10.1016/j.canlet.2008.03.052.

69. Phillip CJ, Giardina CK, Bilir B, Cutler DJ, Lai YH, Kucuk O, et al. Genistein cooperates with the histone deacetylase inhibitor vorinostat to induce cell death in prostate cancer cells. BMC Cancer. 2012;12:145. doi:10.1186/1471-2407-12-145.

70. Verkerk R, Schreiner M, Krumbein A, Ciska E, Holst B, Rowland I, et al. Glucosinolates in Brassica vegetables: the influence of the food supply chain on intake, bioavailability and human health. Mol Nutr Food Res. 2009;53 Suppl 2:S219. doi:10.1002/mnfr.200800065.

71. Watson GW, Beaver LM, Williams DE, Dashwood RH, Ho E. Phytochemicals from cruciferous vegetables, epigenetics, and prostate cancer prevention. AAPS J. 2013;15(4):951–61. doi:10.1208/s12248-013-9504-4.

72. Clarke JD, Dashwood RH, Ho E. Multi-targeted prevention of cancer by sulforaphane. Cancer Lett. 2008;269(2):291–304. doi:10.1016/j.canlet.2008.04.018.

73. Safe S, Papineni S, Chintharlapalli S. Cancer chemotherapy with indole-3-carbinol, bis(3′-indolyl)methane and synthetic analogs. Cancer Lett. 2008;269(2):326–38. doi:10.1016/j.canlet.2008.04.021.

74. Wong CP, Hsu A, Buchanan A, Palomera-Sanchez Z, Beaver LM, Houseman EA, et al. Effects of sulforaphane and 3,3′-diindolylmethane on genome-wide promoter methylation in normal prostate epithelial cells and prostate cancer cells. PLoS ONE. 2014;9(1):e86787. doi:10.1371/journal.pone.0086787.

75. Khan N, Mukhtar H. Multitargeted therapy of cancer by green tea polyphenols. Cancer Lett. 2008;269(2):269–80. doi:10.1016/j.canlet.2008.04.014.

76. Yang CS, Wang X, Lu G, Picinich SC. Cancer prevention by tea: animal studies, molecular mechanisms and human relevance. Nat Rev Cancer. 2009;9(6):429–39. doi:10.1038/nrc2641.

77. Gupta S, Hastak K, Ahmad N, Lewin JS, Mukhtar H. Inhibition of prostate carcinogenesis in TRAMP mice by oral infusion of green tea polyphenols. Proc Natl Acad Sci U S A. 2001;98(18):10350–5. doi:10.1073/pnas.171326098.

78. Caporali A, Davalli P, Astancolle S, D'Arca D, Brausi M, Bettuzzi S, et al. The chemopreventive action of catechins in the TRAMP mouse model of prostate carcinogenesis is accompanied by clusterin over-expression. Carcinogenesis. 2004;25(11):2217–24. doi:10.1093/carcin/bgh235.

79. Adhami VM, Siddiqui IA, Sarfaraz S, Khwaja SI, Hafeez BB, Ahmad N, et al. Effective prostate cancer chemopreventive intervention with green tea polyphenols in the TRAMP model depends on the stage of the disease. Clin Cancer Res. 2009;15(6):1947–53. doi:10.1158/1078-0432.CCR-08-2332.

80. Kim SJ, Amankwah E, Connors S, Park HY, Rincon M, Cornnell H, et al. Safety and chemopreventive effect of Polyphenon E in preventing early and metastatic progression of prostate cancer in TRAMP mice. Cancer Prev Res (Phila). 2014;7(4):435–44. doi:10.1158/1940-6207.CAPR-13-0427-T.

81. Morey Kinney SR, Zhang W, Pascual M, Greally JM, Gillard BM, Karasik E, et al. Lack of evidence for green tea polyphenols as DNA methylation inhibitors in murine prostate. Cancer Prev Res (Phila). 2009;2(12):1065–75. doi:10.1158/1940-6207.CAPR-09-0010.

82. Hovestadt V, Jones DT, Picelli S, Wang W, Kool M, Northcott PA, et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. Nature. 2014;510(7506):537–41. doi:10.1038/nature13268.

83. UCSC Genome Browser. UCSC genome bioinformatics available from: http://genome.ucsc.edu/.

84. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12(6):996–1006. doi:10.1101/gr.229102.

85. Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, et al. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. PLoS ONE. 2013;8(12):e81148. doi:10.1371/journal.pone.0081148.

86. MethBase. MethBase: a reference methylome database. Available from: http://smithlabresearch.org/software/methbase/.

87. Li Y, Kong D, Ahmad A, Bao B, Dyson G, Sarkar FH. Epigenetic deregulation of miR-29a and miR-1256 by isoflavone contributes to the inhibition of prostate cancer cell growth and invasion. Epigenetics. 2012;7(8):940–9. doi:10.4161/epi.21236.